Private Approximation of the Set Cover Problem : Extended Abstract(Theory of Computer Science and Its Applications)

AUTHOR(S):

Yashiro, Masatoshi; Tanaka, Keisuke

CITATION:

Yashiro, Masatoshi ...[et al]. Private Approximation of the Set Cover Problem : Extended Abstract(Theory of Computer Science and Its Applications). 数理解析研究所講究録 2007, 1554: 48-55

ISSUE DATE:

2007-05

URL:

http://hdl.handle.net/2433/80973

数理解析研究所講究録
第1554巻 2007年 48-55

集合被覆問題に関する近似アルゴリズムの秘匿性
Private Approximation of the Set Cover Problem (Extended Abstract)

八代 正俊* 田中 圭介*
Masatoshi Yashiro Keisuke Tanaka

東京工業大学 数理・計算科学専攻
Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology

Abstract— Private approximation, introduced by Feigenbaum, Ishai, Malkin, Nissim, Strauss, and Wright, allows us to find approximate solutions with disclosing as little information as possible. In STOC 2006, Beimel, Carmi, Nissim, and Weinreb studied the private approximation for both the vertex cover and the max exact 3SAT problems. In this paper, we consider the set cover problem where the costs of all sets are polynomially bounded. We show that there exists neither a deterministic nor a randomized private approximation. We also consider the case that the frequencies of all elements are equal. We show that in this case there exist no deterministic private approximation.

Keywords: private approximation, set cover problem.

1 Introduction

Private approximation is an algorithm which is more efficient than exact computation and that maintain the privacy of the data, that is, the output of private approximation does not leak any information of the input.

Feigenbaum, Ishai, Malkin, Nissim, Strauss, and Wright [?] introduced the notion of the private approximation of functions. Roughly speaking, an approximation function $\hat{g}$ is called private approximation with respect to the target function g, if $\hat{g}(x)$ reveals no more information about x than $g(x)$ does. More formally, there exists a probabilistic polynomial time simulator $\mathcal{M}$ such that the distribution of the simulation output $\mathcal{M}(g(x))$ is indistinguishable from $\hat{g}(x)$. They proposed a function (two-party protocol) which is the private approximation with respect to that for computing the hamming distance between two binary vectors. They also proposed the private approximations of several natural $\#P$-hard problems. After [?], several private approximations were proposed [?, ?, ?].

Approximation algorithm is currently one of the main research fields in computer science. The design of algorithms for approximating NP-hard problems has attracted substantial attention in the last few decades, as did the research on proving hardness of approximation. Halevi, Krauthbamer, Kushilevitz, and Nissim [?] discussed the private approximation of NP-hard problems. They proved that there exists no private approximation for computing the size of minimum vertex cover within approximation ratio $n^{1-\epsilon}$. Their proof used the *sliding-window reduction* that translates a SAT instance ϕ to an instance G of the vertex cover problem. If ϕ is satisfiable then G has the vertex cover of size z, otherwise any vertex cover for G is of size at least $z+1$. The definition of the private approximation in [?] is almost the same as that by Feigenbaum, Ishai, Malkin, Nissim, Strauss, and Wright [?].

Beimel, Nissim, Carmi, and Weinreb [?] studied the private approximation of both the vertex cover and the max exact 3SAT problems, and Beimel, Hallak, and Nissim [?] studied the private approximation of both the vertex cover and the clustering problem. In order to consider search problems, Beimel et al. [?] proposed a definition of the private approximation which is different from that in [?]. In their definition, an algorithm $\mathcal{A}$ is a private approximation with respect to a privacy structure $\mathcal{R}$, which is an equivalent relation, if the outputs of executing $\mathcal{A}$ on two $\mathcal{R}$-equivalent inputs are computational indistinguishable. Under their definition, they showed that there exists neither a deterministic nor a randomized private approximation of the search problem for a minimum vertex cover within

* Supported in part by NTT Information Sharing Platform Laboratories and Grant-in-Aid for Scientific Research, Ministry of Education, Culture, Sports, Science, and Technology, 16092206.

approximation ratio $n^{1-\epsilon}$.

In this paper, we consider the private approximation of the set cover problem. The vertex cover problem which studied by [?, ?] is essentially the special case of the set cover problem where the frequency of all elements are equal to 2. Therefore, by the result of Beimel et al. [?], we can see there exists no private approximation of the set cover problem where the frequency of all elements are equal to 2. However, in the other case, that is, the frequencies of some elements are not equal to 2, it is not clear whether there exists private approximation of the set cover problem.

In this paper we consider the private approximation of the set cover problem. In the previous paper [?, ?], only the vertex cover problem whose costs of all not fixed that vertices are fixed. In particular, we consider the set cover problem where the costs of all sets are polynomially bounded. We show that there exists neither a deterministic nor a randomized private approximation. We also consider the case that the frequencies of all elements are equal. We show that in this case there exist no deterministic private approximation.

Due to lack of space, the proofs are omitted from this paper. See the full version [?].

2 The Set Cover Problem

In this section, we describe the set cover problem and the frequency.

Definition 2.1 (Set Cover Problem). *Let U be a set of m elements, $\mathcal{S} = \{S_1, \ldots, S_n\}$ a collection of subsets of U, and c: $\mathcal{S} \to \mathbb{Q}^+$ a cost function. We say the set $C \in \{1, \ldots, n\}$ of indices is called a cover of U if the collection of S_i $(i \in C)$ covers all elements in U, that is, $\bigcup_{i \in C} S_i = U$. Given $\langle U, \mathcal{S}, c \rangle$, the set cover problem is to find a minimum cost cover of U.*

Usually, the size of the instance $\langle U, \mathcal{S}, c \rangle$ of the set cover problem is considered as the number of the elements in U. In this paper, we consider the size of the instance of the set cover problem as the number of sets in $\mathcal{S}$. Therefore, the number of the elements in each set in $\mathcal{S}$ is restricted to polynomial of the number of the sets in $\mathcal{S}$.

In this paper, we consider "*polynomial-cost set cover problem*" where the cost of each set is polynomial in the problem size. Let "**Set Cover**" be a polynomial-cost set cover problem.

Definition 2.2 (Frequency). *We define the frequency of an element to be the number of sets the element is in. A useful parameter is the frequency of the most frequent element. Let us denote this by f.*

We call the problem where all elements in U have the equivalent frequency as "*set cover problem with fixed frequency*" .

3 The Approximation and the Private Algorithm

First, we describe the definition of the approximation. The following definition of the approximation can be applied to *minimization* problems. The definition for maximization problems is similar.

Definition 3.1 (Approximation of the Search Problem). *Let g be a function, $\mathcal{A}$ an algorithm for a search problem, and c a cost function. We say that $\mathcal{A}$ is an α-approximation of g if it runs in polynomial time and for all input x,*

$$\sum_{y \in \mathcal{A}(x)} c(y) \leq \alpha \sum_{y \in g(x)} c(y).$$

Next, we describe the definition of the private algorithm, following [?]. We describe the privacy structure which is necessary to define the private algorithm.

Definition 3.2 (Privacy Structure). *A privacy structure $\mathcal{R} \subseteq \{0,1\}^* \times \{0,1\}^*$ is an equivalence relation on instances. For $\langle x, y \rangle \in \mathcal{R}$, we use the notation $x \equiv_{\mathcal{R}} y$.*

We only discuss on the privacy structures of the form $\mathcal{R} = \cup_{n \in \mathbf{N}} \mathcal{R}_n$, where $\mathcal{R}_n$ is an equivalence relation among the instances of size n, such as $\mathcal{S}$ with n sets.

We now define the private algorithm. We say that an algorithm $\mathcal{A}$ is private with respect to a privacy structure $\mathcal{R}$ if the results of executing $\mathcal{A}$ on two $\mathcal{R}$-equivalent inputs are computationally indistinguishable.

Definition 3.3 (Private Algorithm). *Let $\mathcal{R}$ be a privacy structure. A probabilistic polynomial-time algorithm $\mathcal{A}$ is private with respect to $\mathcal{R}$ if for every polynomial-time algorithm $\mathcal{D}$ and for every positive polynomial $p(\cdot)$, there exists some $n_0 \in \mathbf{N}$ such that for every $x, y \in \{0,1\}^*$, $x \equiv_{\mathcal{R}} y$, and $|x| = |y| \geq n_0$,*

$$\left|\Pr[\mathcal{D}(\mathcal{A}(x), x, y) = 1] - \Pr[\mathcal{D}(\mathcal{A}(y), x, y) = 1]\right| \leq \frac{1}{p(|x|)}.$$

That is, when $x \equiv_{\mathcal{R}} y$, any algorithm $\mathcal{D}$ cannot distinguish if the input of $\mathcal{A}$ is x or y.

Next, in order to define the private approximation of the search problem, we define the privacy structure, following [?]. We can regard the decision and the search problems as follows by using the bivariate relation.

$\langle U_1, \mathcal{S}_1, c_1\rangle$	
$U_1 = \{e_1, e_2, e_3, e_4\}$	
$\mathcal{S}_1 = \{S_1, S_2, S_3, S_4\}$	
$S_1 = \{e_1, e_2\}$	$c_1(S_1) = 4$
$S_2 = \{e_2, e_3\}$	$c_1(S_2) = 2$
$S_3 = \{e_3, e_4\}$	$c_1(S_3) = 1$
$S_4 = \{e_1, e_3\}$	$c_1(S_4) = 2$

$\langle U_2, \mathcal{S}_2, c_2\rangle$	
$U_2 = \{e_1, e_2, e_3, e_4\}$	
$\mathcal{S}_2 = \{S_1, S_2, S_3, S_4\}$	
$S_1 = \{e_1, e_2, e_3\}$	$c_2(S_1) = 4$
$S_2 = \{e_2, e_3\}$	$c_2(S_2) = 3$
$S_3 = \{e_4\}$	$c_2(S_3) = 2$
$S_4 = \{e_1, e_3\}$	$c_2(S_4) = 1$

Figure 1: $\langle U_1, \mathcal{S}_1, c_1\rangle \equiv_{\mathcal{R}_{\text{minSC}}} \langle U_2, \mathcal{S}_2, c_2\rangle$. Note that $|\mathcal{S}_1| = |\mathcal{S}_2|$. Both solutions of $\langle U_1, \mathcal{S}_1, c_1\rangle$ and $\langle U_2, \mathcal{S}_2, c_2\rangle$ are equivalent ($\{1, 3\}$ and $\{2, 3, 4\}$).

Definition 3.4. *A bivariate relation Q is polynomially bounded if there exists a constant c such that $|w| \leq |x|^c$ for every $\langle x, w\rangle \in Q$. The decision problem for Q is, given an input x, to decide if there exists an element w such that $\langle x, w\rangle \in Q$ or not. The search problem for Q is, given an input x, to find an element w such that $\langle x, w\rangle \in Q$ if such w exists.*

We now define the privacy structure of the search problem. We require that if two input values have the same set of answers of the search problem, the approximation algorithm should not be able to distinguish between them.

Definition 3.5 (Privacy Structure of the Search Problem). *The privacy structure $\mathcal{R}_Q$ related to the relation Q is defined as follows: $x \equiv_{\mathcal{R}_Q} y$ iff*

- $|x| = |y|$,
- $\langle x, w\rangle \in Q$ *iff* $\langle y, w\rangle \in Q$ *for every* w.

That is, $x \equiv_{\mathcal{R}_Q} y$ if they have the same set of solutions.

Finally, we give two relations of the problems considered in this paper.

Definition 3.6. *Let* minSC be the minimum set cover relation for `Set Cover`, that is, $\langle\langle U, \mathcal{S}, c\rangle, C\rangle \in$ minSC if C is the minimum cost cover for $\langle U, \mathcal{S}, c\rangle$. In this case, the privacy structure $\mathcal{R}_{\text{minSC}}$ contains all pairs $(\langle U_1, \mathcal{S}_1, c_1\rangle, \langle U_2, \mathcal{S}_2, c_2\rangle)$ where every minimum cost cover $C \in \mathcal{S}$ for $\langle U_1, \mathcal{S}_1, c_1\rangle$ is that for $\langle U_2, \mathcal{S}_2, c_2\rangle$ and vice verse. Similarly, let $\langle U, \mathcal{S}, c\rangle$ be the minimum cost cover relation for `Set Cover` with fixed-frequency.

In Figure ??, we give an example for the relation minSC.

4 Private Approximation of Set Cover

In this section, we show that there exists no deterministic private approximation algorithm of `Set Cover`.

4.1 Definitions

In this section, we describe some definitions.

First, we describe the definition of the private approximation of `Set Cover`.

Definition 4.1 (Private Approximation of the Set Cover Problem). *An algorithm $\mathcal{A}$ is a private α-approximation algorithm for* minSC if:

- $\mathcal{A}$ is a α-approximation algorithm for minSC, and
- $\mathcal{A}$ is private with respect to $\mathcal{R}_{\text{minSC}}$.

In order to analyze the private approximation of the vertex cover problem, Beimel et el. [?] employed "critical vertices" and "relevant vertices". We also employ the notion of "critical" and "relevant" for `Set Cover`.

Definition 4.2 (Critical Set and Relevant Set). *Let U be a set of m elements, $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$ a collection of sets and c a cost function of $\mathcal{S}$. We say that S_i is critical for $\langle U, \mathcal{S}, c\rangle$ if every minimum set cover of $\langle U, \mathcal{S}, c\rangle$ contains S_i. We say that S_i is relevant for $\langle U, \mathcal{S}, c\rangle$ if there exists at least one minimum set cover of $\langle U, \mathcal{S}, c\rangle$ that contains S_i.*

Next, we present the problem related to Definition ??.

Definition 4.3 (The Relevant Set / Non-Critical Set Problem).

Input: a Set U, a collection $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$ of sets, and a cost function c.

Output: "S_i is relevant for $\langle U, \mathcal{S}, c\rangle$" or "$S_i$ is non-critical for $\langle U, \mathcal{S}, c\rangle$".

We next define two special set cover problems. When we construct the algorithm for the Relevant / Non-Critical Set problem in Section ??, they are helpful.

Definition 4.4 ($\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$). *Let U be a set that contains m elements $e_1, \ldots, e_m$, $\mathcal{S} = \{S_1, \ldots, S_n\}$ a collection of subsets of U, c a cost function $\mathcal{S} \to \mathbb{Q}^+$, I a collection of empty sets. For $\langle U, \mathcal{S}, c\rangle$ and for any $S_u \in I$ and $S_t \in \mathcal{S}$, we define $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ as follows.*

The collection $\mathcal{S}^2$ of sets is defined as $\mathcal{S}^2 = \{S_1, \ldots, S_{2n}\} \cup I$ where $S_{i+n} := \{e_{k+m} \mid e_k \in S_i\}$. The set U^2 is defined as $U^2 = \{e_1, \ldots, e_{2n}\}$. The function c^2 is defined as $c^2(S_i) = c(S_i)$ (for $1 \leq i \leq n$), $c^2(S_i) = c(S_{i-l})$ (for $l+1 \leq i \leq 2n$), and $c^2(S_i) = 1$ (for $S_i \in I$).

The collection $\mathcal{S}_{(t,u)}$ of sets is defined as $\mathcal{S}_{(t,u)} = \{S_1, \ldots, S_n\} \cup I$ where $S_t = S_t \cup \{e^*, e^{**}\}$, $S_u = S_u \cup \{e^*\}$, and $S_{u+n} = S_{u+n} \cup \{e^{**}\}$ for some

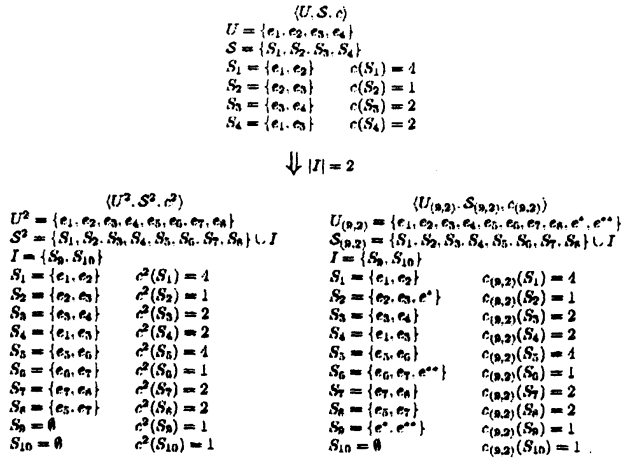

Figure 2: The constructions of $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ $(i = 9,\ j = 2)$ using $\langle U, \mathcal{S}, c\rangle$ and I with size 2.

$e^*, e^{**} \notin U^2$. *The set* $U_{(t,u)}$ *is defined as* $U_{(t,u)} = U^2 \cup \{e^*, e^{**}\}$, *and let* $c_{(t,u)} = c^2$.

We give concrete examples $\mathcal{S}^2$ and $\mathcal{S}_{(t,u)}$ in Figure ??.

4.2 Proofs

In this section, we show that there exists no private approximation algorithm of Set Cover with respect to $\mathcal{R}_{\text{minSC}}$ if $P \neq NP$.

Theorem 4.5. *Let* $\epsilon > 0$ *be a constant and* f *a frequency. If* $P \neq NP$, *then there is no deterministic private* f^ϵ*-approximation algorithm of the search problem of* minSC .

This proof is similar to that in [?]. The outline of the proof is as follows :

1. We construct a Relevant or Non-Critical for Set Cover algorithm from the private approximation algorithm $\mathcal{A}$ with respect to $\mathcal{R}_{\text{minSC}}$.
2. We construct a greedy algorithm that efficiently solves the NP-hard problem from the Relevant or Non-Critical for Set Cover algorithm.
3. If $P \neq NP$, this is a contradiction. Thus there is no private approximation algorithm $\mathcal{A}$ for Set Cover with respect to $\mathcal{R}_{\text{minSC}}$.

In Algorithm ??, we describe a greedy algorithm of Set Cover given an access to the algorithm which decides relevant or non-critical (we call this algorithm Relevant or Non-Critical for Set Cover). We will show that the algorithm Relevant or Non-Critical for Set Cover can be constructed by using oracle access to private approximation algorithms of Set Cover later on.

Algorithm. 1 (Greedy Minimum Set Cover)

Input: a collection of sets $\mathcal{S} = \{S_1, \ldots, S_n\}$, a cost function $c : \mathcal{S} \to \mathbb{Q}^+$, and a set U of m elements.

1. Set $C_s = \emptyset$.
2. If $U = \emptyset$ return C_s.
3. Pick a set $S_i \in \mathcal{S}$ and execute the algorithm Relevant or Non-Critical for Set Cover on $\langle U, \mathcal{S}, c\rangle$ and S_i.
4. If the answer is "Relevant",
   (a) Delete all the elements included in S_i from both U and sets S_j in $\mathcal{S}$.
   (b) $\mathcal{S} \leftarrow \mathcal{S} \setminus \{S_i\}$.
   (c) $C_s \leftarrow C_s \cup \{i\}$.
   (d) Go to STEP2.
5. If the answer is "Non-Critical",
   (a) $\mathcal{S} \leftarrow \mathcal{S} \setminus \{S_i\}$.
   (b) Go to STEP2.

The following claim shows the correctness of the greedy algorithm.

Claim 4.6. *If the algorithm* Relevant or Non-Critical for Set Cover *is polynomial and correct then the algorithm* Greedy Minimum Set Cover *is polynomial and correct.*

We next construct the Relevant or Non-Critical for Set Cover algorithm from a private approximation algorithm for minSC. We adopt the idea of [?].

Claim 4.7. *Let* $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$ *be a collection of sets,* c *a cost function* $\mathcal{S} \to \mathbb{Q}^+$, e^* *an element such that* $e^* \notin U$. *We choose* $i, j \in \{1, \ldots, n\}$ *arbitrary* $i \neq j$, *and define* $\mathcal{S}^* = \{S_1^*, \ldots, S_n^*\}$ *where* $S_i^* = S_i \cup \{e*\}$, $S_j^* = S_j \cup \{e*\}$, *and* $S_k^* = S_k$ *for* $k \neq i, j$. *We also define* $U^* = U \cup \{e^*\}$ *and* $c^*(S_i^*) = c(S_i^*)$.

Then, If S_j *is critical for* $\langle U, \mathcal{S}, c\rangle$, *then* $\langle U, \mathcal{S}, c\rangle \equiv_{\mathcal{R}_{\text{minSC}}} \langle U^*, \mathcal{S}^*, c^*\rangle$.

We can prove the two claims with respect to $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ defined in Definition ??. We use these claims for the proof of the correctness of Algorithm ??.

Claim 4.8. *If* S_u *is critical for* $\langle U, \mathcal{S}, c\rangle$, *then* $\langle U^2, \mathcal{S}^2, c^2\rangle \equiv_{\mathcal{R}_{\text{minSC}}} \langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$.

Claim 4.9. *If* S_u *is not relevant for* $\langle U, \mathcal{S}, c\rangle$, *then* S_t *is critical for* $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$.

Next, by using a private f^ϵ-approximation algorithm we describe the Relevant or Non-Critical for

Set Cover algorithm in Algorithm ??.

Algorithm. 2 (Relevant or Non-Critical for Set Cover)

Input: $(\langle U, \mathcal{S}, c\rangle, S_u)$

1. Let I be a set of $2df^\epsilon + 1$ empty sets. (where $d = \sum_i c(S_i)$)
2. Construct the collection of sets $\mathcal{S}^2$ from $\mathcal{S}$ and I.
3. Execute $\mathcal{A}$ on $\langle U^2, \mathcal{S}^2, c^2\rangle$ and get the output W^2 of $\mathcal{A}$.
4. Choose any set $S_t \in I \setminus W^2$.
5. Construct the collection of sets $\mathcal{S}_{(t,u)}$ from $\mathcal{S}$, I, S_t, and S_u.
6. Execute $\mathcal{A}$ on $\mathcal{S}_{(t,u)}$, and get the output $W_{(t,u)}$ of $\mathcal{A}$.
7. If $W^2 \neq W_{(t,u)}$, return "Non-Critical". Else return "Relevant".

We can show the following claim.

Claim 4.10. *Let $\mathcal{A}$ be a deterministic private approximation algorithm for* minSC, *U a set of m elements, $\mathcal{S} = \{S_1, \ldots, S_n\}$ a collection of subsets of U, and c a cost function, and denote $\mathcal{A}(\langle U, \mathcal{S}, c\rangle)$ be a cover of $\langle U, \mathcal{S}, c\rangle$ that corresponding to indices outputted by $\mathcal{A}$. Then for any set $S_i \in \mathcal{S} \setminus W$, the set S_i is not critical for $\langle U, \mathcal{S}, c\rangle$.*

We must prove Algorithm ?? is correct and running time is polynomial. We can prove the correctness by proving the following two claims. In the proofs of the following claims, we use Claim ??.

Claim 4.11. *If $W^2 \neq W_{(t,u)}$, then S_u is not critical for $\langle U, \mathcal{S}, c\rangle$.*

Claim 4.12. *If $W^2 = W_{(t,u)}$, then S_u is relevant.*

Finally, we show that there is a set chosen in STEP 4 of Algorithm ??.

Claim 4.13. *Let $\epsilon \geq 0$. There is a set $S_i \in I$ such that $S_i \in I \setminus W^2$.*

Therefore we proved Theorem ??.

Remark 4.14. *Since d is polynomial in the size of problem, the sizes of $\mathcal{S}^2$ and $\mathcal{S}_{(t,u)}$ are polynomial in the size of problem. Therefore we prove Theorem ??. When d is exponentially large, the sizes of $\mathcal{S}^2$ and $\mathcal{S}_{(t,u)}$ are are exponentially large. Therefore in this case, we do not get an impossibility result by applying our strategy.*

5 Randomized Private Approximation of the Set Cover Problem

In this section, we show that there exists no randomized private approximation algorithm of Set Cover. The outline of the proof is similar to that in [?]. We execute the approximation algorithm k times to decide whether the set is Relevant or Non-Critical. We prove several claims than correspond to those in Section ??.

We use the Algorithm ??, and we use Algorithm ?? as Relevant or Non-Critical for Set Cover in ??.

Algorithm. 3 (Randomized Relevant or Non-Critical for Set Cover)

Input: $(\langle U, \mathcal{S}, c\rangle, S_u)$.

1. If $\mathcal{S}$ contains less than n_2 sets (where $n_2 = \max\{n_0, n_1\}$), then find if S_u is relevant for $\langle U, \mathcal{S}, c\rangle$ or non-critical for $\langle U, \mathcal{S}, c\rangle$ using exhaustive search.
2. Let I be a set $4df^\epsilon + 2$ sets. (where $d = \sum_i c(S_t)$)
3. Construct the family of sets $\langle U^2, \mathcal{S}^2, c^2\rangle$ from $\langle U, \mathcal{S}, c\rangle$ and I as in Definition ??.
4. Execute k times the algorithm $\mathcal{A}$ on $\langle U^2, \mathcal{S}^2, c^2\rangle$.
5. Choose a set $S_t \in I$ such that S_t appears at most $k/2$ times in $\mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$ in the k executions.
6. Construct the family of sets $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ from $\langle U, \mathcal{S}, c\rangle$, I, S_t, and S_u as in Definition ??.
7. Execute k times Algorithm $\mathcal{A}$ on $\mathcal{S}_{(t,u)}$.
8. If $t \in \mathcal{A}(\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle)$ in at least $0.75k$ of the k executions, then return "Non-Critical" Else return "Relevant".

We can show the following claims.

Claim 5.1. *There is a set $S_t \in I$ such that index t appears at most $k/2$ times in $\mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)$ out of the k executions.*

Claim 5.2. *There exists a constant n_1 such that if*

- *$\langle U^2, \mathcal{S}^2, c^2\rangle$ contains at least n_1 sets,*
- *$\Pr\left[t \in \mathcal{A}(\langle U^2, \mathcal{S}^2, c^2\rangle)\right] < 0.55$, and*
- *$\Pr\left[t \in \mathcal{A}(\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle)\right] > 0.6$,*

then S_u is not critical for $\langle U, \mathcal{S}, c\rangle$

Claim 5.3. *If $\Pr[\{t, u\} \cap \mathcal{A}\langle U, \mathcal{S}, c\rangle) = \emptyset] \leq 0.8$ then S_u is relevant for $\langle U, \mathcal{S}, c\rangle$.*

Claim 5.4. *Let $k > \Omega(\log(df^\epsilon))$. Algorithm* Randomized Relevant or Non-Critical for Set Cover *returns the correct answer with probability $1 - 2^{-O(k)}$*

From the above claims, we can prove the following main theorem.

Theorem 5.5. *Let $\epsilon > 0$ be a constant. If $RP \neq NP$, then there is no randomized private f^{ϵ}-approximation algorithm for* `Set Cover`.

proof. By Claim ?? and Claim ??, if there is a randomized private f^{ϵ}-approximation algorithm for `Set Cover`, then there is a randomized algorithm for the exact search problems for minSC. This algorithm is transformed to the algorithm for decision problem of `Set Cover` (given $\langle U, \mathcal{S}, c\rangle$ and $x \in \mathbb{Q}^+$, decide whether there is a cover of cost at most x). Since this problem is NP-complete, it contradicts $RP \neq NP$. □

6 Private Approximation of the Set Cover Problem with the Fixed Frequency

When the frequency of all elements U are 2, the set cover problem is essentially the same as the vertex cover problem. Therefore, in this section, we consider the situation that the frequencies of all $e \in U$ are equal.

We show that there exists no randomized private approximation algorithm of `Set Cover` with the fixed frequency. The strategy of the proof is similar to that in the previous section, however, the construction of the greedy algorithm depends on whether the cost is fixed or not.

First, we describe the greedy algorithm for the case that the cost is not fixed. In this case, it is easy to construct the greedy algorithm.

Algorithm. 4 (Greedy Minimum Set Cover with fixed frequency (cost is not fixed))

Input: a collection of sets $\mathcal{S} = \{S_1, \ldots, S_n\}$, a cost function $c : \mathcal{S} \to \mathbb{Q}^+$, and a set U of m elements.

1. Set $C_s = \emptyset$.
2. If $U = \emptyset$ return C_s.
3. Pick an element $e_i \in U$ and make a list of all sets that include e_i. We define this list as $L = \{L_1, \ldots L_m\}$.
4. Set $c' \leftarrow c$ and $j = 1$.
5. Execute Algorithm Relevant or Non-Critical for Set Cover with fixed frequency on $\langle U, \mathcal{S}, c\rangle$ and $S_{j'}$ (where $S_{j'} = L_j$).
6. If the answer is "Relevant"
   (a) Delete all the elements included in $S_{j'}$ from both U and sets S_i in $\mathcal{S}$.
   (b) $\mathcal{S} \leftarrow \mathcal{S} \setminus \{S_{j'}\}$.
   (c) $c \leftarrow c'$.
   (d) $C_s \leftarrow C_s \cup \{j'\}$.
   (e) Go to STEP 2.
7. If the answer is "Non-Critical"
   (a) $c(S_{j'}) \leftarrow c(S_{j'}) + 1$.
   (b) $j \leftarrow j + 1$ and go to STEP 5.

We can show that this algorithm is polynomial and correct.

Claim 6.1. *If the algorithm* Relevant or Non-Critical for Set Cover with fixed frequency *is polynomial and correct and the cost of each set is not fixed then the algorithm* Greedy Minimum Set Cover with fixed frequency *is polynomial and correct.*

In Algorithm ??, if the cost is fixed, we can not execute STEP 6-a. Therefore we transform from Algorithm ?? to Algorithm ??.

Algorithm. 5 (Greedy Minimum Set Cover with fixed frequency (cost is fixed))

Input: a collection of sets $\mathcal{S} = \{S_1, \ldots, S_n\}$, a cost function $c : \mathcal{S} \to \mathbb{Q}^+$, and a set U of m elements.

1. Set $C_s = \emptyset$.
2. If $U = \emptyset$ return C_s.
3. Pick an element $e_i \in U$ and make a list of all sets that include e_i. We define this list as $L = \{L_1, \ldots, L_l\}$.
4. Set $\mathcal{S}' \leftarrow \mathcal{S}$, and $j = 1$.
5. If $j \leq l$,
   (a) Execute Algorithm Relevant or Non-Critical for Set Cover with fixed frequency on $\langle U, \mathcal{S}, c\rangle$ and $S_{j'}$ (where $S_{j'} = L_j$).
   (b) If the answer is "Relevant"
      i. Delete all the elements included in $S_{j'}$ from both U and all sets S_i in $\mathcal{S}$.
      ii. $\mathcal{S} \leftarrow \mathcal{S}' \setminus \{S_{j'}\}$.
      iii. $C_s \leftarrow C_s \cup \{j'\}$.
      iv. Go to STEP 2.
   (c) If the answer is "Non-Critical"
      i. If $|S_i| = 1$
         A. $j \leftarrow j + 1$ and go to STEP 5.
      ii. Else
         A. Divide S_i into $|S_i|$ $(= k)$ sets $S_{i_1}, \ldots, S_{i_k}$.
         B. $\mathcal{S} \leftarrow (\mathcal{S} \setminus S_i) \cup S_{i_1} \cup \ldots \cup S_{i_k}$.
         C. $j \leftarrow j + 1$ and go to STEP 5.
6. If $j = l$,
   (a) $C_s \leftarrow C_s \cup h$ where $S_h = L_1$.
   (b) $\mathcal{S} \leftarrow \mathcal{S}' \setminus \{S_h\}$.
   (c) Go to STEP 2.

We can show that this algorithm is polynomial and correctness.

Claim 6.2. *If Algorithm* Relevant or Non-Critical for Set Cover with fixed frequency *is polynomial and correct and the cost of each set is fixed, then Algorithm* Greedy Minimum Set Cover with fixed frequency *is polynomial and correct.*

Next we consider the algorithm Relevant or Non-Critical for Set Cover with fixed frequency. This is the Algorithm ?? where $\langle U_{(t,u)}, \mathcal{S}_{(t,u)}, c_{(t,u)}\rangle$ is replaced with $\langle U_{(t,\ u\ |\ 2(f-2))}, \mathcal{S}_{(t,\ u\ |\ 2(f-2))}, c_{(t,\ u\ |\ 2(f-2))}\rangle$. We now define $\langle U_{(t,\ u\ |\ 2(f-2))}, \mathcal{S}_{(t,\ u\ |\ 2(f-2))}, c_{(t,\ u\ |\ 2(f-2))}\rangle$.

Definition 6.3 $(\langle U_{(t,\ u\ |\ 2(f-2))}, \mathcal{S}_{(t,\ u\ |\ 2(f-2))}, c_{(t,\ u\ |\ 2(f-2))}\rangle)$. *$\langle U^2, \mathcal{S}^2, c^2\rangle$ and I are the same as those defined by Definition ??. We choose the $2(f-2)$ elements from I, and which we denote $S_{k_1}, \ldots S_{k_{2(f-2)}}$. The collection $\mathcal{S}_{(t,\ u\ |\ 2(f-2))}$ of sets is defined as $\mathcal{S}_{(t,\ u\ |\ 2(f-2))} = \{S_1, \ldots, S_n\} \cup I$ where $S_t = \{e^*, e^{**}\}$, $S_u = S_u \cup \{e^*\}$, $S_{u+n} = S_{u+n} \cup \{e^{**}\}$, $S_{k_i} = \{e^*\}$ $(1 \leq i \leq f-1)$, and $S_{k_i} = \{e^{**}\}$ $(f \leq i \leq 2(f-2))$ such that $e^*, e^{**} \notin U^2$. The set $U_{(t,\ u\ |\ 2(f-2))}$ defined as $U_{(t,\ u\ |\ 2(f-2))} = U^2 \cup \{e^*, e^{**}\}$, and $c_{(t,\ u\ |\ 2(f-2))} = c^2$.*

We can easily see $\langle U^2, \mathcal{S}^2, c^2\rangle$ and $\langle U_{(t,\ u\ |\ 2(f-2))}, \mathcal{S}_{(t,\ u\ |\ 2(f-2))}, c_{(t,\ u\ |\ 2(f-2))}\rangle$ are `Set Cover` with fixed frequency if $\langle U, \mathcal{S}, c\rangle$ is `Set Cover` with fixed frequency.

Finally, we can prove Claims ??, ??, ??, and ?? in a similar way as those for the proof of Claims ??, ??, ??, and ??, respectively.

The following two claims are used in the proofs of Claim ?? and Claim ??.

Claim 6.4. *If S_u is critical for $\langle U, \mathcal{S}, c\rangle$, then $\langle U, \mathcal{S}, c\rangle \equiv_{\mathcal{R}_{\text{minSCfixed}}} \langle U_{(t,\ u\ |\ 2(f-2))}, \mathcal{S}_{(t,\ u\ |\ 2(f-2))}, c_{(t,\ u\ |\ 2(f-2))}\rangle$.*

Claim 6.5. *If S_u is not relevant for $\langle U, \mathcal{S}, c\rangle$, then S_i is critical for $\langle U_{(t,\ u\ |\ 2(f-2))}, \mathcal{S}_{(t,\ u\ |\ 2(f-2))}, c_{(t,\ u\ |\ 2(f-2))}\rangle$.*

The following two claims guaranteed the correctness of Algorithm ??.

Claim 6.6. *If $W^2 \neq W_{(t,u\ |\ k_1,\ldots,k_{2(f-2)})}$, then S_u is not critical for $\langle U, \mathcal{S}, c\rangle$.*

Claim 6.7. *If $W^2 = W_{(t,u\ |\ k_1,\ldots,k_{2(f-2)})}$, then S_u is relevant.*

We can prove the following theorem from the above claims and Claim ??.

Theorem 6.8. *Let $\epsilon > 0$ be a constant and f a frequency. If $P \neq NP$, then there is no deterministic private f^ϵ-approximation algorithm of the search problem for* minSCfixed.

7 Concluding Remarks

In this paper, we have considered the set cover problem where the costs of all sets are polynomially bounded. We have shown that there exists neither a deterministic nor a randomized private approximation. We have also considered the case that the frequencies of all elements are equal. We have shown that in this case there exist no deterministic private approximation.

In this paper, we have proved only when the size of the problem is defined as the number of the sets. It might be interesting to consider the problem where the size of the problem is defined as the number of elements. It might be also interesting to consider whether NP-hard problems other than the set cover problem have the private approximation algorithms or not.

Halevi et al. [?] discussed the leakage of information about the approximation algorithms for the minimum set cover problem. Beimel et al. [?] also discussed that for the vertex cover and the exact 3SAT problems. It might be interesting to consider the leakage of information about the approximation algorithms for the minimum set cover problem.

References

[1] A. Beimel, P. Carmi, K. Nissim, and E. Weinreb. Private Approximation of Search Problems. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC 2006)*, pages 119–128, Seattle, WA, USA,, May 2006. ACM Press.

[2] A. Beimel, R. Hallak, and K. Nissim. Private Approximation of Clustering and Vertex Cover. To appear in Fourth Theory of Cryptography Conference - TCC 2007.

[3] J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. Strauss, and R. N. Wright. Secure Multiparty Computation of Approximations. In F. Orejas, P. G. Spirakis, and J. van Leeuwen, editors, *28th International Colloquium, ICALP 2001*, volume 2076, pages 927–938, Crete, Greece, July 2001. Springer-Verlag.

[4] M. J. Freedman, K. Nissim, and B. Pinkas. Efficient Private Matching and Set Intersection. In C. Cachin and J. Camenisch, editors, *Advances in Cryptology - EUROCRYPT 2004*, volume 3027 of *Lecture Notes in Computer Science*, Interlaken, Switzerland, May 2004. Springer-Verlag.

[5] S. Halevi, R. Krauthgamer, E. Kushilevitz, and K. Nissim. Private Approximation of NP-hard Functions. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC 2001)*, pages 550–559, Heraklion, Crete, Greece, July 2001. ACM Press.

[6] P. Indyk and D. Woodruff. Polylogarithmic Private Approximations and Efficient Matching. In T. R. Shai Halevi, editor, *Third Theory of Cryptography Conference - TCC 2006*, volume 3876 of *Lecture Notes in Computer Science*, New York, NY, USA, March 2006. Springer-Verlag.

[7] E. Kiltz, G. Leander, and J. Malone-Lee. Secure Computation of the Mean and Related Statistics. In J. Kilian, editor, *Third Theory of Cryptography Conference - TCC 2005*, volume 3378 of *Lecture Notes in Computer Science*, Cambridge, MA, USA, February 2005. Springer-Verlag.

[8] M. Yashiro and K. Tanaka. Private Approximation of the Set Cover Problem. Technical Report C-234, Dept. of Mathematical and Computing Science, 2006. http://www.is.titech.ac.jp/reserch/research-report/.